Speech Prosody 2016
31 May - 3 Jun 2106, Boston, USA

Speaking rate consistency and variability in spontaneous speech by native and non-native speakers of English

Tuuli Morrill,[1] *Melissa Baese-Berk,*[2] *Ann Bradlow*[3]

[1] George Mason University, Fairfax, Virginia, U.S.A.
[2] University of Oregon, Eugene, Oregon, U.S.A.
[3] Northwestern University, Evanston, Illinois, U.S.A.
tmorrill@gmu.edu, mbaesebe@uoregon.edu, abradlow@northwestern.edu

Abstract

The suprasegmental characteristics of non-native speech are generally described as differing from those of native speech. Recent work has shown that in addition to speaking at overall slower rates, non-native speakers in a reading task are more variable than native speakers in their speaking rate across utterances [1]. However, read speech may contain sources of variability that are specific to processing difficulties associated with reading. In the present study, we examined speaking rate in spontaneous utterances by native speakers of Korean and Mandarin speaking English, and compared them to spontaneous utterances of native English speakers. We measured mean speaking rate within utterances, as well as the amount of rate change (slowing or speeding up of speaking rate) from utterance to utterance. Results indicate that spontaneous speech exhibits the opposite pattern of read speech; non-native speakers are *less* variable than native speakers in spontaneous speech. These findings are attributed to factors involving speaking style and language proficiency, and have implications for understanding the role of suprasegmental variability in speech perception and production.

Index Terms: speaking rate, non-native speech production, variability

1. Introduction

Non-native speech patterns vary from native speech patterns in multiple ways. Most investigations of second language speech production have focused on specific phonemic, phonetic and articulatory patterns that differ from native speakers. These differences are attributed to discrepancies between the native and non-native phoneme inventories or phonotactic constraints [2,3]. However, non-native speech also differs at the supra-segmental level. These differences are described as an effect of transfer between the prosodic structures of the native and non-native languages. In second language production, non-native speakers may differ from native speakers in the placement of stressed syllables, or the production of non-native phrasal patterning [4]. Recent investigations have examined speaking rate, another suprasegmental characteristic, in non-native speech. Overall, non-native speakers have been shown to speak more slowly than native speakers [5]. Recently, non-native speakers have also been shown to exhibit more changes in speaking rate than native speakers [1]. The current study focuses on this recent finding that speaking rate variability differs between native and non-native speakers, and aims to clarify the nature of these differences.

The examination of speaking rate in non-native speech is particularly important because speaking rate is a suprasegmental property of every language. Regardless of differences in the phonological or prosodic structures of a speaker's native and non-native languages, they will constantly produce utterances in which speaking rate can be measured and compared. In addition, the rate at which a person speaks interacts with the production of segments, syllables, and prosodic units, highlighting its potentially significant role in language proficiency. The finding that non-native speech is overall slower than native speech can in part be attributed to lower proficiency in the non-native language. However, recent work has shown that a person's speaking rate in their non-native language is correlated with speaking rate in their native language; a person who speaks relatively slowly will do so in both languages [6,7]. This suggests that the relationship between speaking rate and language proficiency is not entirely straightforward. In other words, in addition to language proficiency, individual differences and possibly speech style, may affect speaking rate.

Recent work by Baese-Berk and Morrill [1] demonstrated that in addition to speaking more slowly, non-native speakers exhibited more variability in speaking rate from utterance to utterance than native speakers did. While most research examining native and non-native speech, including work on speaking rate, has only measured mean differences across populations, Baese-Berk and Morrill examined speaking rate *changes* from utterance to utterance. Comparisons of native Korean, Mandarin and English speakers reading a text in English revealed that non-native speakers changed their rate from utterance to utterance more than native speakers did; in other words, they were more variable in their speaking rate. However, the reasons for increased variability in read speech are not entirely clear. In read speech, vocabulary knowledge and reading skills themselves could affect rate consistency. For example, listeners may slow down or hesitate when reading an unfamiliar word or difficult phrase, and then speed up in the following utterance.

The current study investigates the same non-native speakers' productions in spontaneous speech, where speakers are allowed to choose their vocabulary and phrase structures. In the present study, we ask whether within-speaker variability, in addition to mean differences, characterizes

 doi: 10.21437/SpeechProsody.2016-230

distinctions found between non-native and native spontaneous speech as it did in reading. To do this, we examine speaking rate in spontaneous productions by both native and non-native speakers of English, and measure the amount of change in speaking rate from utterance to utterance. Because non-native speakers have been shown to exhibit greater amounts of change in speaking rate than native speakers in reading, they may also show greater amounts of change in spontaneous speech. On the other hand, if changes in speaking rate in spontaneous speech are influenced by factors other than reading-related processing difficulties, native and non-native speakers may exhibit similar degrees of rate change, or show other distinctive patterns in speaking rate variability.

2. Methods

2.1. Materials

This study used recordings collected for the Archive of L1 and L2 Scripted and Spontaneous Transcripts and recordings (ALLSSTAR), available from The Online Speech/Corpora Archive and Analysis Resource (OSCAAR, https://oscaar.ci.northwestern.edu/index.html). We analyzed data from native Korean speakers (n=9), native Mandarin speakers (n=8), and native English speakers (n=10) in the corpus; these identical speakers' read speech was previously analysed [1]. All non-native speakers were graduate students in the U.S. with high levels of English proficiency. In the current study we analyzed data from two different spontaneous speech tasks: (1) a picture book "story-telling" task, (2) and an interview-style question and answer task. In the story-telling task, speakers described simple scenes that they were viewing as pictures, similar to a children's storybook. In the question and answer task, speakers were answering questions about their lives and describing events of their own choosing. Thus, the story-telling task was relatively structured and may elicit similar vocabulary across speakers, whereas the question and answer task was less structured.

2.2. Analysis

Each speaker's productions were divided into utterances, defined as periods of speech between audible pauses or non-speech sounds (e.g., throat clearing) of 250ms or greater. A script in Praat [8] was used to automatically identify the utterances, as well as syllable nuclei (based primarily on amplitude) (script by de Jong and Wempe, 2008). Pauses were not included in the analysis.

We calculated three measures for each utterance: utterance duration, speaking rate, and rate change across utterances. Utterance duration was simply the temporal duration of each utterance (in seconds). Speaking rate was calculated as the average number of syllables per second produced within each utterance. Rate change across utterances was calculated as it was in Baese-Berk and Morrill's study [1] – the difference between the speaking rate of a given utterance and the speaking rate of the previous utterance. Thus, this value was positive if an utterance was faster than the preceding utterance and negative if an utterance was slower than the preceding utterance. We also calculated the absolute value of the rate change.

3. Results

3.1. Story-telling Task

3.1.1. Speaking Rate and Utterance Duration

Overall, native English speakers spoke at slightly faster rates (M = 4.3 syll/s, SD = 1.1) than the native Mandarin (M = 4.0 syll/s, SD = 0.97) or native Korean speakers (M = 3.7 syll/s, SD = 0.94) in the Story-telling task (Figure 1). All data were analyzed with linear mixed effects regression models using the lme4 package in R [9]. Models were fitted with contrast coding to compare native speakers (English) vs. non-native speakers (Mandarin and Korean together) and to compare Mandarin vs. Korean speakers. A model predicting speaking rate with native language contrasts as fixed factors, and random intercepts for participants showed that the difference between native and non-native speakers was a significant predictor of speaking rate (model R^2 = .18, Type II Wald χ^2 = 16.64, p < .001), i.e, the non-native speakers were significantly slower. This finding replicates previous results suggesting that non-native speakers tend to speak more slowly than native speakers [1,5]. There was no significant difference between the native Mandarin and Korean speaking groups (model R^2 = .18, Type II Wald χ^2 = 1.9677, p = .16).

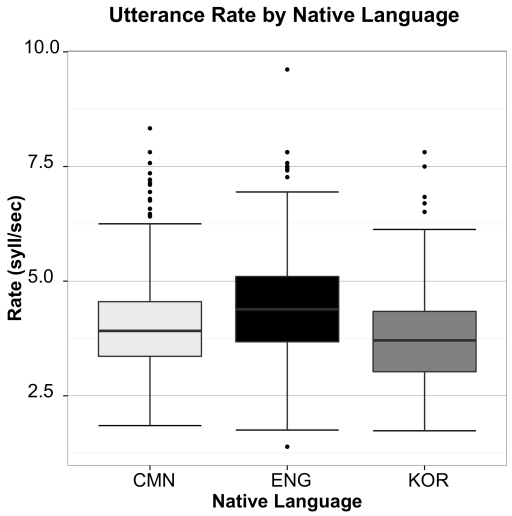

Figure 1: *Medians and quartiles for speaking rate of each utterance produced by native Mandarin (CMN), English (ENG) and Korean (KOR) speakers in the Story-telling task.*

Although the average duration of utterances was slightly longer for native English speakers than the non-native speakers (Figure 2), utterance duration across all groups exhibited a high degree of variability and none of the native language contrasts were a significant predictor of duration (model R^2 = .13, Type II Wald χ^2 < 2.7, p > 0.1).

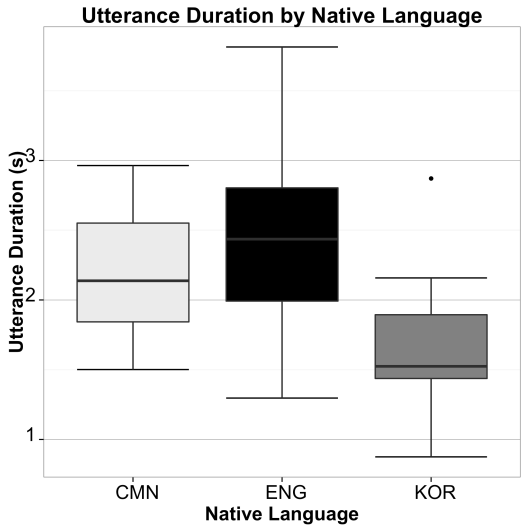

Figure 2: *Medians and quartiles for utterance duration of each utterance produced by native Mandarin (CMN), English (ENG) and Korean (KOR) speakers in the story-telling task.*

3.1.2. Rate Change across utterances

The measure of rate change (the difference between the speaking rate of an utterance and that preceding it) indicated that native speakers of English exhibited slightly greater amounts of rate change from utterance to utterance than non-native speakers in the Story-telling task (Figure 3). This indicates that native English speakers were slowing down and speeding up more than the non-native speakers. A linear mixed effects regression model predicting Rate Change was fit with Native Language contrasts (Native vs. Non-native and Mandarin vs. Korean) and Speaking Rate (syll/s) as fixed factors with interaction terms, and random intercepts for participants (Table1).

Table . *Linear mixed effects regression model predicting Rate Change across utterances in the Story-telling task. Native Language contrasts and Speaking Rate were fixed effects with interaction terms and Participants were random effects.*

	Estimate	Std. Err.	*t* value
Intercept	-3.63	0.15	-24.84
Native vs. Non-Native	**1.04**	**0.41**	**2.55**
Mand. vs. Korean	0.10	0.36	0.27
Rate	**0.89**	**0.03**	**27.65**
Nat. vs. NN*Rate	-0.08	0.08	-0.90
Korean*Rate	0.029	0.08	0.35

Model comparison showed that speaking rate was a significant predictor of rate change (χ^2 = 557.25, p < .001), and that native language contrasts were a significant predictor of rate change (χ^2 = 15.133, p < .01). The difference between native and non-native speakers in predicting rate change was significant (χ^2 = 13.62, p < .01), meaning that native English speakers exhibited more rate change than non-native speakers. There was no difference between the native Mandarin and Korean speaking groups (χ^2 = 1.78, p = .41). In addition, the interaction between speaking rate and native language did not contribute significantly to model fit (χ^2 = 0.96, p = .62), suggesting that the effect of native language on rate change is independent of the overall differences in speaking rate of the utterances.

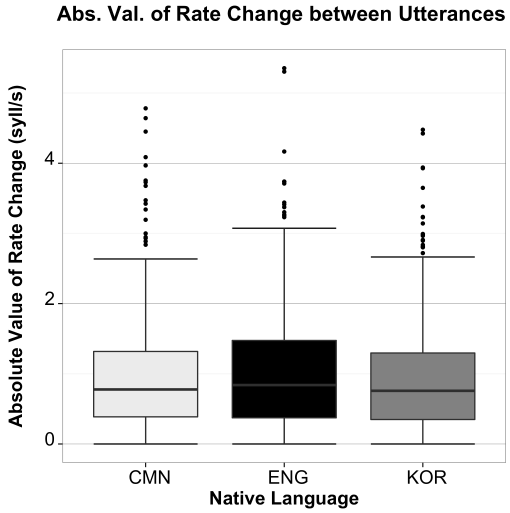

Figure 3: *Medians and quartiles for the amount of rate change from utterance to utterance produced by native Mandarin (CMN), English (ENG) and Korean (KOR) speakers in the story-telling task.*

3.2. Question and Answer task

The speaking rate, utterance duration, and rate change measures in the Question and Answer task followed the same pattern as in the Story-telling task. Once again, native English speakers produced utterances at faster rates (M = 4.4 syll/s, SD = 1.02) than the native Mandarin (M = 3.9 syll/s, SD = 0.94) and Korean speakers (M = 3.7 syll/s, SD = 1.0). Mixed effects models were fitted in R with contrast coding to compare native speakers (English) vs. non-native speakers (Mandarin and Korean together) and to compare Mandarin vs. Korean speakers. A model predicting speaking rate with native language contrasts as fixed factors, and random intercepts for participants confirmed that the difference between native and non-native speakers was a significant predictor of speaking rate (model R^2 = .21, Type II Wald χ^2 = 15.47, p < .001), i.e, the non-native speakers were significantly slower. There was no significant difference between the native Mandarin and Korean speaking groups (model R^2 = .21, Type II Wald χ^2 = 1.55, p = .21). As in the Story-telling task, native English speakers produced the longest utterances on average; here, the contrast between the native and non-native speakers did contribute significantly to model fit (model R^2 = .12, Type II Wald χ^2 = 5.55, p < .05).

A linear mixed effects regression model predicting Rate Change in the question and answer task was fit with Native Language contrasts (Native vs. Non-native) and Speaking Rate (syll/s) as fixed factors with interaction terms, and random intercepts for speakers (Table 2). Again, model comparison showed that speaking rate was a significant predictor of rate change (χ^2 = 1838.7, p < .001), and that native language was also a significant predictor of rate change (χ^2 = 16.71, p < .01). The difference between native and non-native speakers in

predicting rate change was significant (χ^2 = 14.947, p < .01); again, native English speakers exhibited more rate change than non-native speakers. There was no difference between the native Mandarin and Korean speaking groups (χ^2 = 2.29, p = .32), and the interaction between speaking rate and native language did not contribute significantly to model fit (χ^2 = 1.95, p = .38).

Table 2. *Linear mixed effects regression model predicting Rate Change across utterances in the Question and Answer task. Native Language contrasts and Speaking Rate were fixed effects with interaction terms and Participants were random effects.*

	Estimate	Std. Err.	*t* value
Intercept	-4.02	0.11	-37.11
Native vs. Non-Native	**0.57**	**0.31**	**1.88**
Mand. vs. Korean	0.37	0.27	1.41
Rate	**1.01**	**0.02**	**52.56**
Nat. vs. NN*Rate	0.06	0.05	1.19
Korean*Rate	-0.03	0.05	-0.69

4. Discussion

The pattern of results found for both the Story-telling Task and the Question and Answer Task suggests that native and non-native speakers show different amounts of change in speaking rate from utterance to utterance. Interestingly, the direction of this difference is the opposite of that found by Baese-Berk and Morrill [1] in the investigation of read speech by the same non-native speakers. In read speech, English speakers showed *less* rate change than either the native Mandarin speaking group or the native Korean speaking group; in the current study of spontaneous speech, the native English speakers showed *more* rate change than the non-native speakers. This was true for both the Story-telling Task and the Question and Answer Task.

Overall, the amount of rate change exhibited by non-native speakers in spontaneous speech appears to be similar to the amount of rate change for non-native speakers in read speech [1]; in other words, the difference across these studies appears to be driven by increased variability in the native speakers in spontaneous speech. Crucially, while native speakers were faster and *more* variable in spontaneous speech (in the current study), in read speech the native speakers were faster and *less* variable. Taken together, these studies suggest that it is not the faster rates alone that drive rate change across utterances. In read speech, the non-native speakers had been the slowest and the most variable.

Non-native speakers may have been expected to be more consistent in spontaneous than read speech due to the ability to choose the vocabulary and syntactic structures of each utterance. Vocabulary in the Story-telling task included the specific lexical items pictured, but otherwise speakers created their own sentences. This is in contrast to read speech, in which every speaker produced identical utterances. The freedom to choose vocabulary and create sentences could have resulted in increased fluency for non-native speakers; however, it does not appear to have decreased the amount of rate change from utterance to utterance in comparison to read speech. Although there may be processing difficulties associated with reading that result in rate variability for non-native speakers, the current study demonstrates that there are also sources of variability in spontaneous speech. Additional research will be needed to determine what factors contribute to variability in non-native speech across speaking styles, and whether these factors are similar or different. If the nature of speaking rate variability differs across speech styles, perceptions of fluency of non-native reading vs. spontaneous speech may also differ.

These findings demonstrate a large amount of speaking rate change in spontaneous speech for native speakers of English. Stylistic choices that affect speaking rate (e.g., dramatic affect) may play a role in native-like productions of story-telling speech. While the Story-telling Task in the current study is most closely associated with this style, the Question and Answer task also elicited anecdotal stories from participants' lives, and speakers may have employed story-telling elements in both types of speech. Thus, one question for future research pertains to the relationship between specific speech styles and speaking rate variability.

Another question for future research pertains to the nature of the variability in different speech styles. The variability examined in this study is change in rate from one utterance to the next. However, it is also possible that speakers are varying their rate *within* utterances. For example, they could be speeding up or slowing down in the middle of an utterance – this type of variability would not be captured in the Rate Change measure used here. It is worth noting that some previous studies have found a relationship between utterance duration and speaking rate [10,11]; however, in an examination of these factors in the current study, no clear directionality in this relationship was found. In other words, the longest utterances were not the fastest or the slowest utterances. In addition to the relationship between utterance length and rate, it would be beneficial to examine effects of syllable structure on rate variation – especially considering possible deviations in non-native productions of syllables, known to affect rate metrics [12].

This study also raises questions about the relationship between language proficiency and speaking rate variability. Previous research has shown that speakers exhibit similar speaking rates in their first and second languages [6,7]; the current study raises the question of whether speakers also alter their speaking rate in similar ways in native and non-native productions. These results suggest that speaking rate variability could be an indication of fluency or proficiency in spontaneous speech, whereas it could indicate disfluency in reading. If a certain amount of speaking rate variability is necessary to sound native-like in a second language, highly fluent speakers may exhibit variability in a second language in a similar way to native speakers (e.g., as in the Story-telling task). However, it is also possible that high degrees of variability are associated with initial stages of acquisition or with difficult tasks (e.g., reading). As proficiency increases, speakers may first focus on attaining consistency in speaking rate, and then progress to manipulating rate for different communicative contexts. Future research will examine this possibility.

5. Acknowledgements

We would like to thank Sung Shim Choi, Zhiyan Gao, and Paul Olejarczuk for their assistance with this project, and the Speech Communication Group at Northwestern for access to ALLSSTAR. This work was partially funded by the Program in Linguistics at George Mason University and by a Faculty Research Award from the University of Oregon to MMB.

6. References

[1] Baese-Berk, M. M. & Morrill, T. H. Speaking rate consistency in native and non-native speakers of English. *J. Acoust. Soc. Am.* **138,** EL223–EL228 (2015).

[2] Flege, J. E. Second language speech learning: Theory, findings, and problems. *In Speech perception and linguistic experience: Issues in cross-language speech research.* York Press (1995).

[3] Davidson, L. Phonetic, phonemic, and phonological factors in cross-language discrimination of phonotactic contrasts. *J. Exp. Psychol. Hum. Percept. Perform.* **37,** 270–282 (2011).

[4] Archibald, J. in *Phonological acquisition and phonological theory* (ed. Archibald, J.) 81–109 Lawrence Erlbaum, (1995).

[5] Guion, S. G., Flege, J. E., Liu, S. H. & Yeni-Komshian, G. H. Age of learning effects on the duration of sentences produced in a second language. *Appl. Psycholinguist.* **21,** 205–228 (2000).

[6] Bradlow, A. R. *et al.* Language- and talker-dependent variation in global features of native and non-native speech. *Proc. XVIIth Int. Congr. Phon. Sci. Hong Kong* (2011).

[7] De Jong, N. H., Groenhout, R., Schoonen, R. & Hulstijn, J. H. Second language fluency: Speaking style or proficiency? Correcting measures of second language fluency for first language behavior. *Appl. Psycholinguist.* **36,** 223–243 (2015).

[8] Boersma, P. & Weenink, D. PRAAT: doing phonetics by computer (Version 5.0.21) [Computer program]. (2005). at <http://www.praat.org/>

[9] Bates, D., Maechler, M. & Bolker, B. *lme4: Linear mixed-effects models using S4 classes. R package version 3.2.2*. (2014).

[10] Jacewicz, E., Fox, R. A. & Wei, L. Between-speaker and within-speaker variation in speech tempo of American English. *J. Acoust. Soc. Am.* **128,** 839–850 (2010).

[11] Quené, H. Multilevel modeling of between-speaker and within-speaker variation in spontaneous speech tempo. *J. Acoust. Soc. Am.* **123,** 1104–1113 (2008).

[12] Trouvain, T. & Möbius, B. Sources of variation of articulation rate in native and non-native speech: comparisons of French and German. *Proc. Speech Prosody 7* (2014).